

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 06-052119

(43)Date of publication of application : 25.02.1994

(51)Int.Cl. G06F 15/16

G06F 12/00

G06F 13/00

H04L 12/28

(21)Application number : 04-226475 (71)Applicant : HITACHI LTD

(22)Date of filing : 03.08.1992 (72)Inventor : DOMYO SEIICHI

SAKURABA TATSUTOSHI

KURODA SAWAKI

NAKANO HIROHIKO

KOBAYASHI ATSUSHI

(54) REPLICATION FILE REFERENCING SYSTEM

(57)Abstract:

PURPOSE: To provide a system for efficiently using a replica server instead of a file server in a distributed data processing system.

CONSTITUTION: A client 120 connected to a communication network 152 gives the reference request of a file 112 in a file server 110 through a communication network 151. When a replica 122 is stored in the disk of the client 120, a gate 140 in the mid-way of a path recognizes it and pieces of information specifying 110, 112, 120 and 122 are written

in an update fact table 143. When a client 130 refers to the file 112, the gate 140 in the mid-way of the path recognizes the reference request and alters the reference request to be transmitted to the file server 110 in an original case so that it is to be transmitted to the client 120. The client 120 plays the role of the replica server only to the file 112.

LEGAL STATUS [Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

* NOTICES *

JP0 and NCIP1 are not responsible for any damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.**** shows the word which can not be translated.

3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] Consist of two or more communication networks, and each communication network is equipped with the gate which controls two or more data processors which have a file storing means, and communication between communication networks. It is a rep RIKETO file reference method in the distributed data processing system in which communication between data processors is possible. Said gate The file in which the data

processor in a self-communication network exists in the file server in other communication networks is referred to. It has the updating fact table which had the information which specifies the data processor (replica server) with which said file server and file which show that the rep RIKETO file was stored in the file storing means, and a rep RIKETO file exist, and its rep RIKETO file registered. The communication to the data processor in the other communication networks from the data processor in a self-communication network is supervised. When it investigates whether this file server and the file are registered with reference to said updating fact table when this communication is what refers to the file which exists in a file server, and not registered, Permit communication with this file server and a file server, a file, and the data processor of a communicating agency are recognized from a communication result. When the information which specifies a file server, a file, the data processor (replica server) that is a communicating agency, and its rep RIKETO file is registered into said updating fact table and it registers with it, The rep RIKETO file reference method characterized by being constituted so that the reference place of a file may be changed into the replica server in which the rep RIKETO file of a file exists from said file server and the data processor of a communicating agency can refer to a rep RIKETO file.

[Claim 2] It is the rep RIKETO file reference method characterized by forming the reception place modification authorization flag which permits changing a reception place into said replica server from a file server to said gate into the transmit data at the time of said data processor communicating with a file server in a rep RIKETO file reference method according to claim 1.

[Claim 3] In a rep RIKETO file reference method according to claim 1 said file server According to the file reference demand from a data processor, rep RIKETO file permission data are added to the file information of a communication result. Said data processor creates a rep RIKETO file based on said rep RIKETO file permission data. Said gate The rep RIKETO file reference method characterized by detecting said rep RIKETO file permission data, and updating said updating fact table based on these authorization data by supervising the communication on said communication network.

[Claim 4] In a rep RIKETO file reference method according to claim 3 said file server The data which show an effective reference term to said rep RIKETO file permission data are added. Into the transmit data at the time of communicating with a file server, said data processor forms the newest replica reference flag which shows the reference demand of the newest rep RIKETO file. Said gate Modification time is registered with said effective reference term at the time of the renewal of said updating fact table based

on said authorization data. A reference place is changed into the replica server in which the newest rep RIKETO file exists when said newest replica reference flag is set up into the transmit data of a data processor. The rep RIKETO file reference method with which said effective reference term is characterized by changing a reference place into the replica server in which the longest rep RIKETO file exists when said newest replica reference flag is not set up.

[Claim 5] In a rep RIKETO file reference method according to claim 1 said gate A data delivery means to control communication with the data processor in a self-communication network, and the data processor in other communication networks, A communication monitor means to supervise the contents of said communication, and an updating fact cognitive means by which the commo data sent from said communication monitor means recognizes that it is a thing about a rep RIKETO file, and updates said updating fact table, It recognizes that the commo data sent from said communication monitor means is the file reference demand to a file server. A reference demand cognitive means to pass control to a postscript reference place modification means while applying interruption to said data delivery means and stopping delivery of the commo data to other communication networks, The rep RIKETO file reference method characterized by having a reference place modification means to change the reference place of a file with reference to said updating fact table based on said commo data, and to direct this modification for said data delivery means.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Industrial Application] This invention relates to the suitable rep RIKETO file reference method for employment with the sufficient effectiveness of the rep RIKETO file in a distributed data processing system.

[0002]

[Description of the Prior Art] A distributed data processing system is a system which carries out each processing, connecting two or more data processing system, such as a

personal computer and a workstation, to the communication network called LAN (Local Area Network), and delivering and receiving information through a communication network mutually. There is a distributed file system which enables access of the file which exists in other data processing system from one data processing system on a communication network as an important technique in a distributed data processing system. A distributed file system enables two or more data processing system to share the file which exists in one data processing system. Here, the data processing system which accesses the data processing system with which the file shared exists at the shared file on a file server, and a call and a file server will be called a client.

[0003] If a specific shared file is accessed frequently, a demand may concentrate on the file server and it may become a bottleneck on the engine performance of the whole distributed data processing system. Then, by putting the copy of this file also on the 2nd file server, and processing a part of access request to this file by the 2nd file server, a load distribution is carried out between two file servers, and aiming at the dissolution of the aforementioned bottleneck is performed. Since access to that of the file concerned is not interrupted and recovery also becomes easy by using a copy as an alternative even if file destruction occurs by putting the copy of the same file on two or more data processing system, it also becomes the improvement in dependability of the whole distributed data processing system. It is called replication to create the replica of the file or a rep RIKETO file and a call, and a copy for each of the copy of the same file, and the data processing system which owns the copy of a file is called replica server. a distributed file system -- an A CM and computing Sir BEIZU -- No. 4 (1990) is explained in PP 321-374 (ACM, Computing Surveys, Vol.22, No.4(1990) PP 321-374), and the 22nd volume is discussed by PP 339-340 of this reference about replication.

[0004]

[Problem(s) to be Solved by the Invention] Employment in the network of a global scale where the distributed file system interconnected not LAN but LAN of ** - will be expected from now on. For example, it uses in order to refer to the file in the office in the U.S. from the office in Tokyo or Osaka. The use of replication then mentioned above becomes important, when controlling concentration of the overhead of the file server in a worldwide scale, and the increment in traffic. If it says in the upper example, the client of Tokyo and Osaka will examine how to access a duplex in the U.S., and how a U.S. replica server is put on Tokyo and the client in Japan accesses a replica server. Compared with the former, generally the time amount which requires latter one for a communication link is also short, and does not have the overhead of a U.S. file server, either.

[0005] However, it is actually a problem where a replica server is put. Conventionally, the server only for replication was installed about the file server which exists in a long distance (U.S.). The client needed to have the step which was conscious of the whereabouts of a dedicated server, and refer to the replica for it. Moreover, concentration of access to the replica server by two or more clients will be inevitably large in the overhead of processing of a replica server, and will increase the traffic of a communication network. The technical problem of this invention is to offer the rep RIKETO file reference method carried out as [refer to / instead of a file server / the replica in a replica server], in case the client which referred to the file in a long-distance file server first is made into the replica server of the file and other clients refer to the file in a file server.

[0006]

[Means for Solving the Problem] In order to attain the above-mentioned purpose, this invention prepares an updating fact table in the gate which controls communication between communication networks, and the information which specifies in the data processor (a replica server) with which said file server and file which shows having stored the rep RIKETO file in the file storing means with reference to the file to which the data processor in a self-communication network exists in the file server in other communication networks in this updating fact table, and a rep RIKETO file exist, and its rep RIKETO file is registered. Said gate supervises the communication to the data processor in the other communication networks from the data processor in a self-communication network, and when it is what refers to the file in which this communication exists in a file server, it investigates whether this file server and the file are registered with reference to said updating fact table. And when not registered, communication with this file server is permitted, a file server, a file, and the data processor of a communicating agency are recognized from a communication result, and the information which specifies a file server, a file, the data processor (replica server) that is a communicating agency, and its rep RIKETO file as said updating fact table is registered.

[0007] When registered, the reference place of a file is changed into the replica server in which the rep RIKETO file of a file exists from said file server, and the data processor of a communicating agency enables it to refer to a rep RIKETO file.

[0008]

[Function] Since the reference demand place of a client is changed into a replica server from a file server when there is a replica server which stores the rep RIKETO file of the file to which a client exists in the file server which requires reference, it can

communicate at a high speed far and efficient file reference is attained rather than it communicates with a file server. Moreover, concentration of network traffic and the burden of a server decrease by not having a specific exclusive replica server. Furthermore, when two or more rep RIKETO files of the same file exist by adding the information on the expiration date of a rep RIKETO file, reference of the optimal rep RIKETO file is made easy.

[0009]

[Example] Hereafter, the example of this invention is explained using drawing. Drawing 1 is the block diagram having shown the outline of the distributed data processing system which can apply this invention. It is the data processing system which became independent, respectively, and each data processing system is connected to the common communication networks 150 and 152, and a communication link and data transfer are possible for the clients 120 and 130 which are the data processing system which accesses the shared file on the file server 110 which is the data processing system with which the file shared exists, and a file server to mutual through a wide area network 151 in between the gate 140 and the gate 142. Each gate consists of data processing system, and the communication fact table 143 is connected to the gate 140. For example, a file server 110 and clients 120 and 130 are highly efficient workstations (WS), the gate 140 is the exchange in which high-speed memory access is possible, the updating fact table of 143 is on memory, and the example of a typical configuration uses a Local Area Network (LAN) as a communication network. The file server 110 and the client 120 have disk units 111 and 121 for file storing, respectively, and refer to the files 112, 113, and 122 stored there for them from a client 130.

[0010] The specific file 112 is considered. First, a client 120 requires reference of a file 112 of a server 110. A server 110 permits creating a replica file to a client 120. A client 120 stores the copy of a file 112 in a disk unit 121 as a replica file 122. The client in which a replica file exists is called replica server. The client 120 in which the replica file 122 exists becomes a replica server. The data which permit a replica during this activity of a series of pass through the gate 140. Said replica authorization data are added to the copy body of a file 112, or before a copy, are separate mail and are transmitted to a client 120 from a server 110. When the gate 140 supervised and discovers the replica authorization data which transmit a network top, it stores in its table 143 the updating fact that the replica of the file 112 of a server 110 was created under management of a client 120.

[0011] Then, when a client 130 requires reference of a file 112 of a server 110, the gate 140 recognizes the demand fact of transmitting a communication network, and searches

said updating fact table 143. When the reference place which the demand fact shows, and the updating place which the updating fact shows are in agreement, the reference place of a client 130 is changed into the replica server 120 from a server 110. Therefore, the replica server 120 processes the file reference demand of a client 130. A client 130 will access a replica 122 instead of a file 112.

[0012] Drawing 2 is the conventional block diagram. It is the big difference from drawing 1 that the replica server 220 of dedication exists. Among drawing 2, although a client 230 and the replica server 220 are on the same communication network 252, it exists in a separate network in fact in many cases, and a client 230 does not know existence of the replica server 220 in many cases. Therefore, a client 230 goes the replica 221 in the replica server 220 to not knowing to refer to [of the file 211 in a file server 120].

[0013] In this example, it is the description that a client becomes the replica server of each file without placing the replica server of dedication. In drawing 1, another file 113 is stored in the disk 131 of a client as a replica 133. Therefore, clients 120 and 130 serve as a replica server of a file 112, and a replica server of a file 113, respectively. The gate 140 takes charge of the processing which changes a file reference place into a replica server at the time of demand passage that clients 120 and 130 should perform only processing which requires the file reference to a file server 110. The communication link through a communication network is performed by capturing to the protocol which defined the communication procedure of a transmitting side and a receiving side, and the specification method of the data processor which should be received, i.e., the specification method of the address, includes it in this. It is carried out by the data transfer's itself generating an electric vibration on a channel, or detecting it, and each data processing system reads the address information contained in the detected electric oscillation, and a communication link is materialized by judging whether it is the communication link to oneself.

[0014] Drawing 3 shows the configuration of the gate 140. The conventional gate 240 has only a data delivery means 300 to send the data between a communication network 250 and 252 mutually. The gate 140 of this example has the reference place modification means means 301, the updating fact cognitive means 302, the reference demand cognitive means 303, the communication monitor means 304, the updating fact table 143, and connectors 310 and 311 independently of the data delivery means 300. It is a processor with separate means 300-304, and a multiprocessor configuration is taken and each processor operates under a separate program. As for the updating fact table 143, it is desirable that it is data which exist on the memory which means 301-303

access in common, after data processing's accelerating. In addition, without making means 300-303 into a separate processor, you may constitute from one processor and it is not necessary to say that it is good also as a multiprocessor configuration of the small number.

[0015] Drawing 4 is one example of the updating fact table 143. It constitutes from a column 401 which shows an updating agency (server), a column 402 which shows an updating place (client), and a column 403 which shows the expiration date specified by an updating place. For example, the data of the character string with which the data of a column 401 and a column 402 combined the network address and file name of data processing system, and a column 403 are an integer which shows time of day at the time of the global standard.

[0016] Drawing 5 shows procedure when the gate 140 receives the commo data of renewal of a replica. Steps 501-503 are procedures which send commo data, and are the procedures in which steps 510-511 recognize update information. Drawing 6 shows procedure when the gate 140 receives the commo data of file reference. It is the procedure in which steps 601-604 send commo data, and steps 610-615 are the procedures of changing the reference place of commo data.

[0017] Hereafter, according to this example, actuation of the gate 140 is explained using drawing 1 and drawing 3 - drawing 6. First, actuation in case a client 120 refers to a file 112 to a server 110 is explained. The connector 310 which is an interface with a communication network 152 processes the communicative destination address other than electrical installation. That is, when the appointed address is the address of network 150 direction, commo data is incorporated and the processing is made to carry out to the data delivery means 300 (step 601). The data delivery means 300 sets commo data by the protocol of a wide area network 151 (step 602), and transmits commo data from a connector 311 (step 603).

[0018] The commo data of a file 112 and the data in which replica authorization is shown are answered by the data delivery means 300 through a connector 311 (step 501). The data delivery means 300 sets commo data by the protocol of a communication network 152 (step 502), and transmits it from a connector 310 (step 503). In that case, the communication monitor means 304 supervises the signal which a connector 311 sends to the data delivery means 300 (step 510), and transmits a signal to the updating fact cognitive means 302. When commo data has recognized it as the data which a replica updates, the updating fact cognitive means 302 is a format as shown in drawing 4, and stores the updating fact in the updating fact table of 143 (step 511).

[0019] Below, actuation in case a client 130 refers to a file 112 to a server 110 is

explained. A reference demand passes through the gate 140 through a connector 310, the data delivery means 300, and a connector 311 (steps 601-603). The communication monitor means 304 supervises the signal which a connector 310 sends to the data delivery means 300, and transmits a signal to the reference demand cognitive means 303 (step 610). The reference demand cognitive means 303 interprets a signal, if it gets to know that it is the file reference demand to a file server, it will search the updating fact table 143 (step 611), and it checks whether the same file already exists in a communication network 152 as a replica. Processing is ended when the updating fact does not exist in the updating fact table 143. When the updating fact exists in the updating fact table 143 (step 612), the reference demand cognitive means 303 interrupts the data delivery means 300, and applies 615 (step 613), and it stops that the data delivery means 300 sends a file reference demand to a communication network 151 through a connector 311 (step 604). The reference demand cognitive means 303 passes control after a termination to the reference place modification means 301. The reference place modification means 301 changes the reference place of data based on the updating fact of the updating fact table 143 (step 614). The data delivery means 300 sends a reference demand to the replica server of a modification place (step 616), and the replica server 120 of a modification place sends reference data to a client 130 (step 617).

[0020] Next, the data transfer after modification of the reference place of the data shown at step 614 of drawing 6, step 616, and step 617 and modification is explained using the example of the transmit data 700 from the client 130 shown in drawing 7 to a file server 110, and the received data 710 obtained as a result. The reference demand 700 contains the flag 704 (for example, code of a bit string) which permits that the data division 703 which show the reference demand of a header for the address 702 grade of the address 701 of a transmitting agency or a reception place to communicate and the specific file 111, and the gate 140 with the updating fact table 143 may change a reception place. Moreover, a flag 705 is the newest replica reference flag (about this, it mentions later). The communication monitor means 304 determines to perform processing of henceforth for referring to the replica server 120 instead of a file server 110, when the flag 704 of a reference demand is set up. If processing (steps 611-614) for referring to the replica server 120 will be performed and a reference place will be changed, the reference place modification means 301 will be overwritten to the address 701 of a reception place with which the client 130 specified the data of the address of a reception place in which the replica server 120 is shown. The data delivery means 300 and a reference demand are sent to the replica server 120 of a modification place instead of a file server 110. 710 contains the data division 713 which show the address

711 of a client 130, the address 712 of the replica server 120, and the contents of the file server, and the flag 714 which shows that it is data from not a file server but the replica server 120 as a result of the reference demand 700 which the replica server 120 sends to a client 130.

[0021] In the another example of step 614 of drawing 6 of this invention, it has an updating fact table 143 like drawing 9, and as shown in drawing 8, in case two or more rep RIKETO files of the same file exist, the procedure of referring to the replica (the newest replica) which passed through the gate 140 at the end, or a replica (the longest effective replica) with the longest expiration date of reference is offered. The updating fact table 143 of drawing 9 shows an example when the original file name 901 which combined Server Name and a file name, the file name 902 which combined replica Server Name and a rep RIKETO file name, the expiration date 903 of the replica notified from the file server, and the time of day 904 which updated the data of items 901-903 when it passed through the gate are included. In drawing 8, whether the newest replica reference flag 705 (for example, bit string) with which the reference place modification means 301 is included in the reference demand 700 which the client 130 transmitted is set up judges (step 801). In case the file name 703 which is in agreement with the file 121 of a reference demand exists, when the newest reference flag 705 has set up, the file updating time of day 904 is searched, the rep RIKETO file 902 of the newest modification time is made into a modification place (step 803), in not setting up, it searches the expiration date 903 of a replica, and the expiration date of reference uses the longest rep RIKETO file 902 as a modification place (step 802).

[0022]

[Effect of the Invention] In the network which interconnects on a world-wide scale, instead of a client referring to the file in a file server by preparing an updating fact table in the gate in the middle of passage, and using this, the rep RIKETO file in a replica server can be referred to, and, according to this invention, there is effectiveness which enables high-speed file reference. Moreover, by not placing a specific exclusive fixed replica server, concentration of traffic is avoided, the load distribution of a distributed file system is promoted, and there is effectiveness which the burden of a server lessens. Furthermore, when two or more rep RIKETO files of the same file exist by adding the information on the expiration date of a rep RIKETO file, there is effectiveness which makes reference of the optimal rep RIKETO file easy.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is the block block diagram showing the outline of the distributed data processing system which is a candidate for application in one example of this invention.

[Drawing 2] It is the block block diagram showing the outline of the conventional distributed data processing system.

[Drawing 3] It is the block diagram showing the configuration of the gate.

[Drawing 4] It is drawing showing the structure of an updating fact table.

[Drawing 5] It is the flow chart which shows the processing in the gate when the gate receives the commo data of renewal of a replica.

[Drawing 6] It is the flow chart which shows the processing which changes the reference place in the gate.

[Drawing 7] It is drawing showing the received data obtained as a result of [its] the transmit data from a client to a file server.

[Drawing 8] It is the flow chart of the replica reference processing at the time of two or more rep RIKETO files of the same file existing.

[Drawing 9] It is drawing showing an example of an updating fact table.

[Description of Notations]

110 File Server

111, 121, 131 Disk

112 113 File

120 130 Client

122 133 Replica file

140 Gate

143 Updating Fact Table

150 152 Network

151 Wide Area Network

300 Data Delivery Means

301 Reference Place Modification Means

302 Updating Fact Cognitive Means

303 Reference Demand Cognitive Means

304 Communication Monitor Means

310 311 Connector

(19)日本国特許庁(JP)

(12)公開特許公報(A)

(11)特許出願公開番号

特開平6-52119

(43)公開日 平成6年(1994)2月25日

(51)Int.Cl. ⁵	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 15/16	3 7 0 M	8840-5L		
12/00	5 4 5 A	8526-5B		
13/00	3 5 7 Z	7368-5B		
H 0 4 L 12/28		8529-5K	H 0 4 L 11/ 00	3 1 0 Z
			審査請求 未請求	請求項の数5(全 12 頁)

(21)出願番号 特願平4-226475

(22)出願日 平成4年(1992)8月3日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 道明 誠一

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所システム開発研究所内

(72)発明者 櫻庭 健年

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所システム開発研究所内

(72)発明者 黒田 澤希

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所システム開発研究所内

(74)代理人 弁理士 笹岡 茂 (外1名)

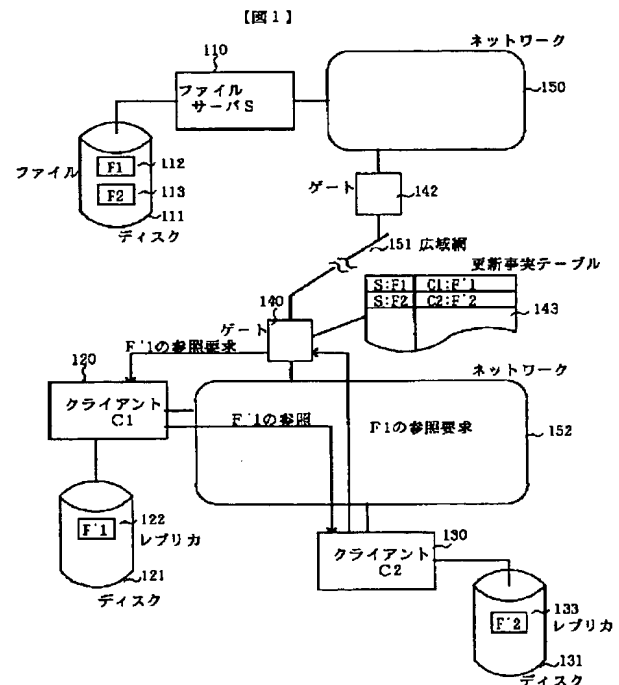
最終頁に続く

(54)【発明の名称】 レプリケートファイル参照方式

(57)【要約】

【目的】 分散データ処理システムにおける、ファイルサーバの代わりにレプリカサーバを効率よく利用する方式を提供する。

【構成】 ファイルサーバ110にあるファイル112の参照要求を通信ネットワーク152に接続されたクライアント120が通信ネットワーク151を通じて行い、クライアント120のディスクにレプリカ122を格納する時、このことを経路途中のゲート140が認知し、上記110、112、120、122を特定する情報を更新事実テーブル143に書き込む。その後、クライアント130が、前記ファイル112の参照を行うときに、経路途中の前記ゲート140が、参照要求を認知し、本来ならば、前記ファイルサーバ110に送達される参照要求を、前記クライアント120へ送達するよう変更する。前記クライアント120は、ファイル112に関してのみ、レプリカサーバの役目をする。



【特許請求の範囲】

【請求項1】 複数の通信ネットワークからなり、各通信ネットワークはファイル格納手段を有する複数のデータ処理装置と通信ネットワーク間の通信を制御するゲートを備え、データ処理装置間での通信が可能な分散データ処理システムにおけるレプリケートファイル参照方式であって、

前記ゲートは、

自通信ネットワーク内のデータ処理装置が他通信ネットワーク内のファイルサーバに存在するファイルを参照してレプリケートファイルをファイル格納手段に格納したことを示す前記ファイルサーバとファイルとレプリケートファイルの存在するデータ処理装置（レプリカサーバ）およびそのレプリケートファイルを特定する情報を登録された更新事実テーブルを備え、

自通信ネットワーク内のデータ処理装置からの他通信ネットワーク内のデータ処理装置への通信を監視し、該通信がファイルサーバに存在するファイルを参照するものであるとき前記更新事実テーブルを参照して該ファイルサーバおよびファイルが登録されているか否かを調べ、登録されていないとき、該ファイルサーバとの通信を許可し、通信結果からファイルサーバとファイルと通信元のデータ処理装置を認知して、前記更新事実テーブルにファイルサーバとファイルと通信元であるデータ処理装置（レプリカサーバ）およびそのレプリケートファイルを特定する情報を登録し、

登録されているとき、ファイルの参照先を前記ファイルサーバからファイルのレプリケートファイルが存在するレプリカサーバに変更し、通信元のデータ処理装置がレプリケートファイルを参照できるよう構成されたことを特徴とするレプリケートファイル参照方式。

【請求項2】 請求項1記載のレプリケートファイル参照方式において、

前記データ処理装置はファイルサーバと通信する際の送信データ中に、受信先をファイルサーバから前記レプリカサーバに変更することを前記ゲートに対して許可する受信先変更許可フラグを設けたことを特徴とするレプリケートファイル参照方式。

【請求項3】 請求項1記載のレプリケートファイル参照方式において、

前記ファイルサーバは、データ処理装置からのファイル参照要求に応じて通信結果のファイル情報にレプリケートファイルの許可データを付加し、

前記データ処理装置は前記レプリケートファイルの許可データに基づきレプリケートファイルを作成し、

前記ゲートは、前記通信ネットワーク上の通信を監視することにより前記レプリケートファイルの許可データを検出し、該許可データに基づき前記更新事実テーブルを更新するようにしたことを特徴とするレプリケートファイル参照方式。

【請求項4】 請求項3記載のレプリケートファイル参照方式において、

前記ファイルサーバは、前記レプリケートファイルの許可データに有効参照期限を示すデータを付加し、

前記データ処理装置はファイルサーバと通信する際の送信データ中に、最新のレプリケートファイルの参照要求を示す最新レプリカ参照フラグを設け、

前記ゲートは、前記許可データに基づく前記更新事実テーブルの更新時に前記有効参照期限と共に更新時刻を登録し、データ処理装置の送信データ中に前記最新レプリカ参照フラグが設定されているときには最新のレプリケートファイルの存在するレプリカサーバに参照先を変更し、前記最新レプリカ参照フラグが設定されていないときには前記有効参照期限が最も長いレプリケートファイルの存在するレプリカサーバに参照先を変更するようにしたことを特徴とするレプリケートファイル参照方式。

【請求項5】 請求項1記載のレプリケートファイル参照方式において、

前記ゲートは、

自通信ネットワーク内のデータ処理装置と他通信ネットワーク内のデータ処理装置との通信を制御するデータ送達手段と、

前記通信の内容を監視する通信監視手段と、

前記通信監視手段から送られた通信データがレプリケートファイルに関するものであることを認知し、前記更新事実テーブルを更新する更新事実認知手段と、

前記通信監視手段から送られた通信データがファイルサーバへのファイル参照要求であることを認知し、前記データ送達手段に割込みをかけ他通信ネットワークへの通信データの送達を中止させると共に後記参照先変更手段に制御を渡す参照要求認知手段と、

前記通信データを基に前記更新事実テーブルを参照してファイルの参照先を変更し、該変更を前記データ送達手段に指示する参照先変更手段を備えることを特徴とするレプリケートファイル参照方式。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は分散データ処理システムにおけるレプリケートファイルの効率の良い運用に好適なレプリケートファイル参照方式に関する。

【0002】

【従来の技術】 分散データ処理システムは、LAN（Local Area Network）と呼ばれる通信ネットワークにパーソナルコンピュータやワークステーションなどのデータ処理システムを複数接続し、相互に通信ネットワークを介して情報の授受を行いながら、それぞれの処理を遂行するシステムである。分散データ処理システムにおける重要な技術として、通信ネットワーク上の1つのデータ処理システムから他のデータ処理システムに存在するファイルのアクセスを可能とする分散

ファイルシステムがある。分散ファイルシステムは、1つのデータ処理システムに存在するファイルを複数のデータ処理システムが共用することを可能とする。ここでは、共用されるファイルが存在するデータ処理システムをファイルサーバと呼び、ファイルサーバ上の共用ファイルにアクセスするデータ処理システムをクライアントと呼ぶことにする。

【0003】特定の共用ファイルが頻繁にアクセスされるとそのファイルサーバに要求が集中し、分散データ処理システム全体の性能上のボトルネックとなる可能性がある。そこでこのファイルのコピーを第2のファイルサーバにも置き、このファイルへのアクセス要求の一部を第2のファイルサーバで処理することにより、2つのファイルサーバの間で負荷分散し、前記のボトルネックの解消を図ることが行われる。同一ファイルのコピーを複数のデータ処理システムに置くことにより、ファイル破壊が発生してもコピーを代替として用いることにより、当該ファイルへのアクセスが中断せず、また回復も容易となるので、分散データ処理システム全体の信頼性向上にもなる。同一ファイルのコピーの各々をそのファイルのレプリカ、あるいはレプリケートファイルと呼び、コピーを作成することをレプリケーションと云い、ファイルのコピーを所有しているデータ処理システムをレプリカサーバと云う。分散ファイルシステムについてはエー・シー・エム、コンピューティング サーベイズ、第22巻、第4号(1990年)PP321-374(ACM, Computing Surveys, Vol. 22, No. 4(1990)PP321-374)において解説されており、レプリケーションについては同文献のPP339-340に論じられている。

【0004】

【発明が解決しようとする課題】分散ファイルシステムは、単一のLANではなく、LANを相互接続した世界的規模のネットワークでの運用が今後期待されている。たとえば、米国のオフィスにあるファイルを東京や大阪のオフィスから参照するために利用する。そのときに前述したレプリケーションの利用は、世界規模でのファイルサーバのオーバーヘッドの集中や、通信量の増加を抑制する上で重要となる。上の例で言えば、東京と大阪とのクライアントが米国に2重にアクセスする方法と、東京に米国のレプリカサーバを置き、日本国内のクライアントは、レプリカサーバをアクセスする方法を検討する。前者にくらべて後者の方が、通信にかかる時間も一般に短く、米国のファイルサーバのオーバーヘッドもない。

【0005】しかし、実際に、レプリカサーバを何処に置くかは問題である。従来は、遠距離(米国)に存在するファイルサーバについて、レプリケーション専用のサーバを設置した。クライアントは、専用サーバの所在を意識したステップを持って、レプリカを参照する必要がある。また、複数のクライアントによるレプリカサー

バへのアクセスの集中は、必然的にレプリカサーバの処理のオーバーヘッドを大きく、通信ネットワークのトラフィックを増すこととなる。本発明の課題は、遠距離のファイルサーバ内のファイルを最初に参照したクライアントをそのファイルのレプリカサーバとし、他のクライアントがファイルサーバ内のファイルを参照する際に、ファイルサーバの代わりにレプリカサーバ内のレプリカを参照できるようにするレプリケートファイル参照方式を提供することにある。

【0006】

【課題を解決するための手段】上記目的を達成するため、本発明は、通信ネットワーク間の交信を制御するゲートに更新事実テーブルを設け、この更新事実テーブルには、自通信ネットワーク内のデータ処理装置が他通信ネットワーク内のファイルサーバに存在するファイルを参照してレプリケートファイルをファイル格納手段に格納したことを示す前記ファイルサーバとファイルとレプリケートファイルの存在するデータ処理装置(レプリカサーバ)およびそのレプリケートファイルを特定する情報が登録される。前記ゲートは、自通信ネットワーク内のデータ処理装置からの他通信ネットワーク内のデータ処理装置への交信を監視し、該交信がファイルサーバに存在するファイルを参照するものであるとき前記更新事実テーブルを参照して該ファイルサーバおよびファイルが登録されているか否かを調べる。そして、登録されていないとき、該ファイルサーバとの交信を許可し、交信結果からファイルサーバとファイルと交信元のデータ処理装置を認知して、前記更新事実テーブルにファイルサーバとファイルと交信元であるデータ処理装置(レプリカサーバ)およびそのレプリケートファイルを特定する情報を登録する。

【0007】登録されているとき、ファイルの参照先を前記ファイルサーバからファイルのレプリケートファイルが存在するレプリカサーバに変更し、交信元のデータ処理装置がレプリケートファイルを参照できるようにしている。

【0008】

【作用】クライアントが参照を要求するファイルサーバに存在するファイルのレプリケートファイルを格納しているレプリカサーバがあるとき、クライアントの参照要求先をファイルサーバからレプリカサーバに変更するため、ファイルサーバと交信するよりはるかに高速に交信することができ、効率的なファイル参照が可能となる。また、特定の専用レプリカサーバを持たないことで、ネットワークのトラフィックの集中や、サーバの負担が少なくなる。さらに、レプリケートファイルの有効期限の情報を付加することにより、同一のファイルの複数のレプリケートファイルが存在するときに、最適なレプリケートファイルの参照を容易にしている。

【0009】

10

20

30

40

50

【実施例】以下、図を用いて本発明の実施例を説明する。図1は本発明が適用可能な分散データ処理システムの概要を示した構成図である。共用されるファイルが存在するデータ処理システムであるファイルサーバ110、ファイルサーバ上の共用ファイルにアクセスするデータ処理システムであるクライアント120、及び130はそれぞれ独立したデータ処理システムであり、各データ処理システムは共通の通信ネットワーク150および152に接続されており、ゲート140とゲート142間を広域網151を介して相互に通信、及びデータ転送が可能である。各ゲートはデータ処理システムからなり、ゲート140には、通信事実テーブル143が接続されている。例えばファイルサーバ110、及びクライアント120、及び130は高性能なワークステーション(WS)であり、ゲート140は、高速なメモリアクセスが可能な交換機であり、メモリ上に143の更新事実テーブルがあり、通信ネットワークとしてローカルエリアネットワーク(LAN)を使用するのが典型的な構成の例である。ファイルサーバ110、及びクライアント120はそれぞれファイル格納のためにディスク装置111、及び121を有しており、そこに格納されたファイル112、113、及び122は、クライアント130から参照することができる。

【0010】特定のファイル112を考える。最初に、ファイル112の参照をサーバ110にクライアント120は要求する。サーバ110は、レプリカファイルを作成することをクライアント120に許可する。クライアント120は、ファイル112のコピーを、ディスク装置121にレプリカファイル122として格納する。レプリカファイルが存在するクライアントをレプリカサーバと云う。レプリカファイル122が存在するクライアント120はレプリカサーバになる。この一連の作業中に、レプリカを許可するデータは、ゲート140を通過する。前記レプリカ許可データは、ファイル112のコピー本体に付加されるか、コピー以前に別便で、サーバ110よりクライアント120へ送信される。ゲート140は、ネットワーク上を転送するレプリカ許可データを監視し、発見した際に、サーバ110のファイル112のレプリカがクライアント120の管理下に作成された更新事実を、自らのテーブル143に格納する。

【0011】その後、ファイル112の参照をサーバ110にクライアント130が要求した際に、ゲート140は、通信ネットワークを伝達する要求事実を認知し、前記更新事実テーブル143を検索する。要求事実が示す参照先と、更新事実が示す更新先が一致するとき、クライアント130の参照先を、サーバ110からレプリカサーバ120へ変更する。よって、クライアント130のファイル参照要求は、レプリカサーバ120が処理する。クライアント130は、ファイル112の代わりにレプリカ122をアクセスすることとなる。

【0012】図2は、従来の構成図である。専用のレプリカサーバ220が存在することが、図1との大きな違いである。図2中、クライアント230とレプリカサーバ220は、同一の通信ネットワーク252上にあるが、実際には別個のネットワークに存在することが多く、クライアント230は、レプリカサーバ220の存在を知らないことが多い。したがって、クライアント230は、レプリカサーバ220内のレプリカ221を知らずに、ファイルサーバ120内のファイル211の参照に行く。

【0013】本実施例においては、専用のレプリカサーバを置かないで、クライアントが個々のファイルのレプリカサーバになることが、特徴である。図1において、別のファイル113は、クライアントのディスク131にレプリカ133として格納している。したがって、クライアント120、及び130は、それぞれファイル112のレプリカサーバ、ファイル113のレプリカサーバとなる。クライアント120、及び130は、ファイルサーバ110へのファイル参照を要求する処理のみを行えばよく、ゲート140が、要求通過時に、ファイル参照先をレプリカサーバへ変更する処理を受け持つ。通信ネットワークを介した通信は送信側、及び受信側の通信手順を定めたプロトコルに乗っ取って行われ、この中には受信すべきデータ処理装置の指定方法、即ちアドレスの指定方法も含んでいる。データ転送自体は通信路上に電気的な振動を発生させ、あるいはそれを検出することにより行われ、通信は各データ処理システムは検出した電気振動に含まれるアドレス情報を読み取って、自らへの通信であるか否かを判定することにより成立する。

【0014】図3は、ゲート140の構成を示したものである。従来のゲート240は、通信ネットワーク250と252間のデータを相互に送達するデータ送達手段300のみを持つ。本実施例のゲート140は、データ送達手段300と独立に、参照先変更手段301、更新事実認知手段302、参照要求認知手段303、通信監視手段304と、更新事実テーブル143と、コネクタ310、311を持つ。手段300~304とは、別々のプロセッサであり、マルチプロセッサ構成を取り、夫々のプロセッサは別々のプログラムの下に動作する。更新事実テーブル143は手段301~303とが共通にアクセスするメモリ上に存在するデータであることが、データ処理の高速化のうで望ましい。なお、手段300~303を、別々のプロセッサとせず、1つのプロセッサで構成してもよく、また、少ない個数のマルチプロセッサ構成としてもよいことは云うまでもない。

【0015】図4は、更新事実テーブル143の一実施例である。更新元(サーバ)を示す欄401と、更新先(クライアント)を示す欄402と、更新先が指定した有効期限を示す欄403とで構成する。たとえば、欄4

01と欄402のデータは、データ処理システムのネットワークアドレスとファイル名とを組合せた文字列、欄403のデータは、世界標準時刻を示す整数である。

【0016】図5は、ゲート140がレプリカ更新の通信データを受けた場合の処理手順を示す。ステップ501～503は、通信データを送達する手順であり、ステップ510～511が、更新情報を認知する手順である。図6は、ゲート140がファイル参照の通信データを受けた場合の処理手順を示す。ステップ601～604が通信データを送達する手順であり、ステップ610～615が、通信データの参照先を変更する手順である。

【0017】以下、図1及び図3～図6を用いて、本実施例にしたがい、ゲート140の動作を説明する。まず、クライアント120がサーバ110へファイル112を参照する場合の動作を説明する。通信ネットワーク152とのインタフェースであるコネクタ310は電気的接続の他に、通信のあて先アドレスの処理を行う。即ち、指定アドレスがネットワーク150方面のアドレスであった場合は、通信データを取り込み、その処理をデータ送達手段300に行わせる（ステップ601）。データ送達手段300は、通信データを広域網151のプロトコルに合わせて（ステップ602）、コネクタ311より通信データを送信する（ステップ603）。

【0018】ファイル112の通信データ、およびレプリカ許可を示すデータが、コネクタ311を介してデータ送達手段300に返信される（ステップ501）。データ送達手段300は、通信データを通信ネットワーク152のプロトコルに合わせて（ステップ502）、コネクタ310より送信する（ステップ503）。そのさいに、交信監視手段304は、コネクタ311がデータ送達手段300へ送達する信号を監視し（ステップ510）、更新事実認知手段302に信号を送信する。更新事実認知手段302は、通信データがレプリカの更新するデータと認知したときに、図4に示すような形式で、143の更新事実テーブルに更新事実を格納する（ステップ511）。

【0019】つぎに、クライアント130がサーバ110へファイル112を参照する場合の動作を説明する。コネクタ310、データ送達手段300、コネクタ311を介して、参照要求がゲート140を通過する（ステップ601～603）。交信監視手段304は、コネクタ310がデータ送達手段300へ送達する信号を監視し、参照要求認知手段303に信号を送信する（ステップ610）。参照要求認知手段303は、信号を解釈し、ファイルサーバへのファイル参照要求であることを知ると、更新事実テーブル143を検索し（ステップ611）、同一のファイルがすでにレプリカとして通信ネットワーク152に存在するかどうかを確認する。更新事実が更新事実テーブル143に存在しない場合は処理を終了する。更新事実が更新事実テーブル143に存在

した場合（ステップ612）には、参照要求認知手段303は、データ送達手段300に割り込み615をかけ（ステップ613）、データ送達手段300がファイル参照要求をコネクタ311を介して、通信ネットワーク151へ送達するのを中止させる（ステップ604）。中止後、参照要求認知手段303は、参照先変更手段301に制御を渡す。参照先変更手段301は、更新事実テーブル143の更新事実をもとに、データの参照先を変更する（ステップ614）。データ送達手段300が変更先のレプリカサーバに参照要求を送達し（ステップ616）、変更先のレプリカサーバ120は、クライアント130へ参照データを送達する（ステップ617）。

【0020】次に、図7に示すクライアント130からファイルサーバ110への送信データ700とその結果得られた受信データ710の実施例を用いて、図6のステップ614、ステップ616、ステップ617で示したデータの参照先の変更と変更後のデータ転送について説明する。参照要求700は、送信元のアドレス701や受信先のアドレス702等の通信するためのヘッダと特定のファイル111の参照要求を示すデータ部703と更新事実テーブル143を持つゲート140が受信先を変更してもよいことを許可するフラグ704（たとえば、ビット列のコード）を含む。また、フラグ705は最新レプリカ参照フラグである（これについては後述する）。交信監視手段304は、参照要求のフラグ704が設定されている場合にファイルサーバ110の代わりにレプリカサーバ120を参照するための以後の処理を行なうことを決定する。レプリカサーバ120を参照するための処理（ステップ611～614）が行なわれ参照先を変更することになると、参照先変更手段301は、レプリカサーバ120を示す受信先のアドレスのデータをクライアント130が指定した受信先のアドレス701に上書きする。データ送達手段300、参照要求をファイルサーバ110ではなく、変更先のレプリカサーバ120に送達する。レプリカサーバ120がクライアント130へ送達する参照要求700の結果710は、クライアント130のアドレス711、レプリカサーバ120のアドレス712、ファイルサーバの内容を示すデータ部713と、ファイルサーバではなくレプリカサーバ120からのデータであることを示すフラグ714を含む。

【0021】本発明の図6のステップ614の別の実施例では、図9のような更新事実テーブル143を備え、図8に示すように、同一ファイルの複数のレプリケートファイルが存在する際に、ゲート140を最後に通過したレプリカ（最新レプリカ）を、あるいは参照の有効期限が最も長いレプリカ（最長有効レプリカ）を参照する手順を提供する。図9の更新事実テーブル143は、サーバ名とファイル名を組合わせたオリジナルなファイル

名901と、レプリカサーバ名とレプリケートファイル名を組合せたファイル名902と、ファイルサーバから通知されたレプリカの有効期限903と、ゲートを通じた際に項目901～903のデータを更新した時刻904を含んだ場合の実施例を示す。図8において、参照先変更手段301が、クライアント130が送信した参照要求700に含まれる最新レプリカ参照フラグ705（たとえば、ビット列）が設定されているかどうかの判定する（ステップ801）。参照要求のファイル121と一致するファイル名703が存在する際に、最新参照フラグ705が設定している場合にはファイル更新時刻904を検索し、最新更新時刻のレプリケートファイル902を変更先とし（ステップ803）、設定しない場合にはレプリカの有効期限903を検索し、参照の有効期限が最も長いレプリケートファイル902を変更先として利用する（ステップ802）。

【0022】

【発明の効果】本発明によれば、世界規模で相互接続するネットワークにおいて、通過途中のゲートに更新事実テーブルを設け、これを利用することにより、クライアントはファイルサーバ内のファイルを参照する代わりに、レプリカサーバ内のレプリケートファイルを参照することができ、高速なファイル参照を可能とする効果がある。また、特定の固定的な専用レプリカサーバを置かないことで、トラフィックの集中を避け、分散ファイルシステムの負荷分散を促進し、サーバの負担が少なくする効果もある。さらに、レプリケートファイルの有効期限の情報を付加することにより、同一のファイルの複数のレプリケートファイルが存在するときに、最適なレプリケートファイルの参照を容易にする効果もある。

【図面の簡単な説明】

【図1】本発明の1実施例における適用対象である分散*

* データ処理システムの概要を示すブロック構成図である。

【図2】従来の分散データ処理システムの概要を示すブロック構成図である。

【図3】ゲートの構成を示すブロック図である。

【図4】更新事実テーブルの構造を示す図である。

【図5】ゲートがレプリカ更新の通信データを受けた場合のゲートにおける処理を示すフローチャートである。

【図6】ゲートにおける参照先を変更する処理を示すフローチャートである。

【図7】クライアントからファイルサーバへの送信データおよびその結果得られた受信データを示す図である。

【図8】同一ファイルの複数のレプリケートファイルが存在する際のレプリカ参照処理のフローチャートである。

【図9】更新事実テーブルの一例を示す図である。

【符号の説明】

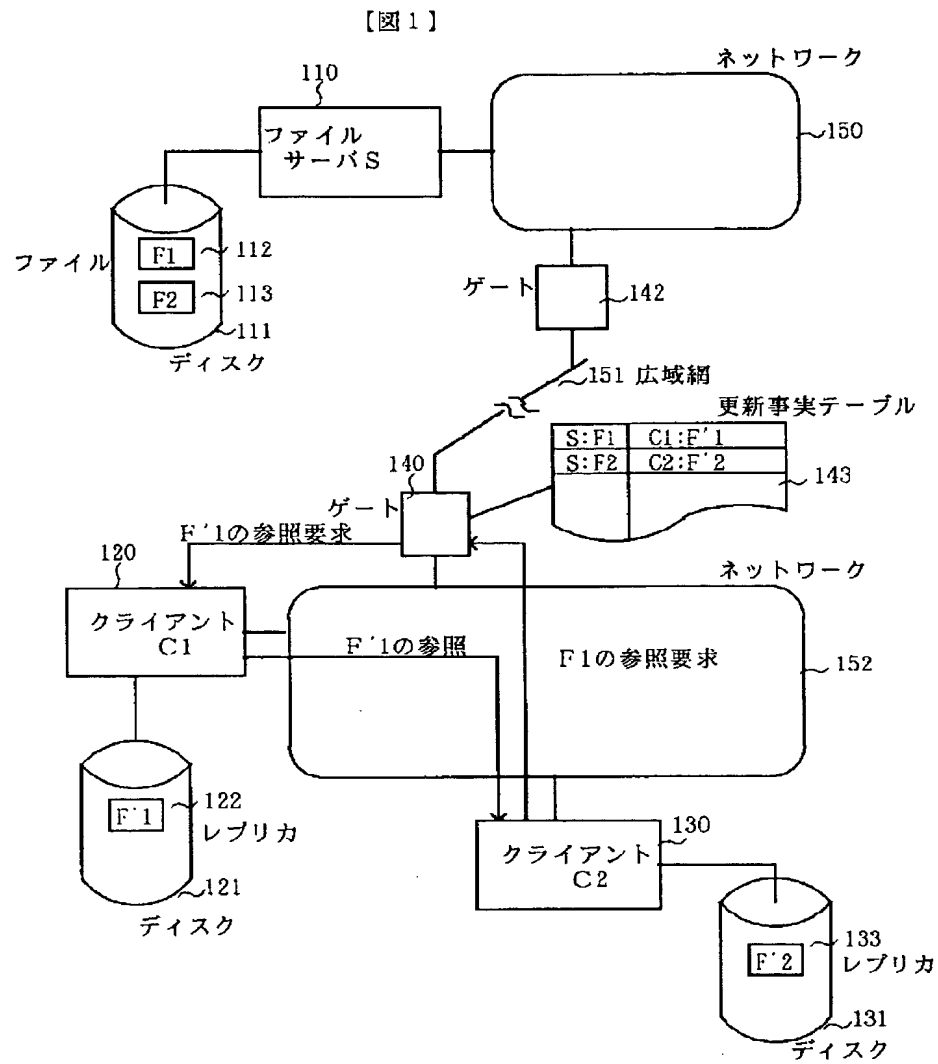
110 ファイルサーバ
111、121、131 ディスク
112、113 ファイル
120、130 クライアント
122、133 レプリカファイル
140 ゲート
143 更新事実テーブル
150、152 ネットワーク
151 広域網
300 データ送達手段
301 参照先変更手段
302 更新事実認知手段
303 参照要求認知手段
304 交信監視手段
310、311 コネクタ

【図4】

【図4】

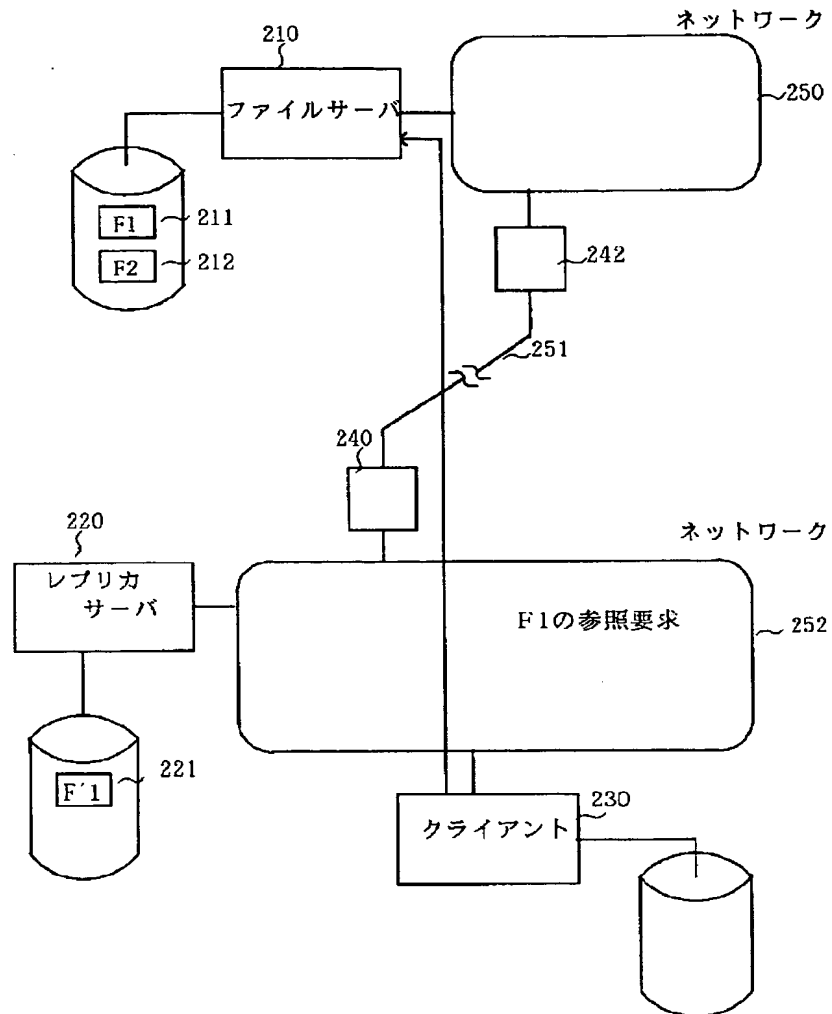
サーバ	レプリカ	有効期限	～143
S : F1	C1 : F'1	t1	
S : F2	C2 : F'2	t2	
⋮	⋮	⋮	

【図1】



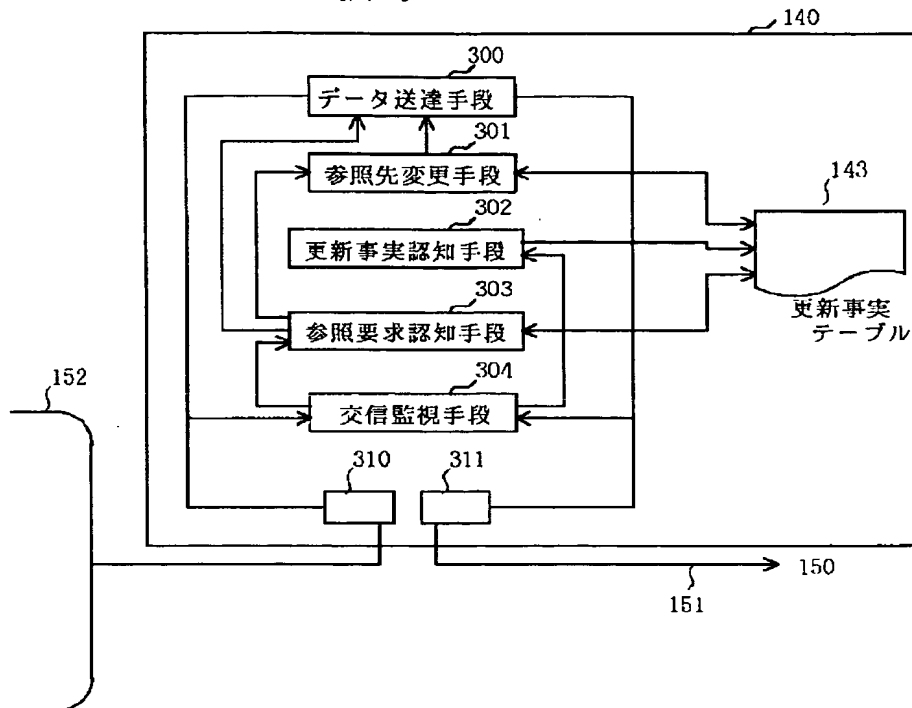
【図2】

【図2】



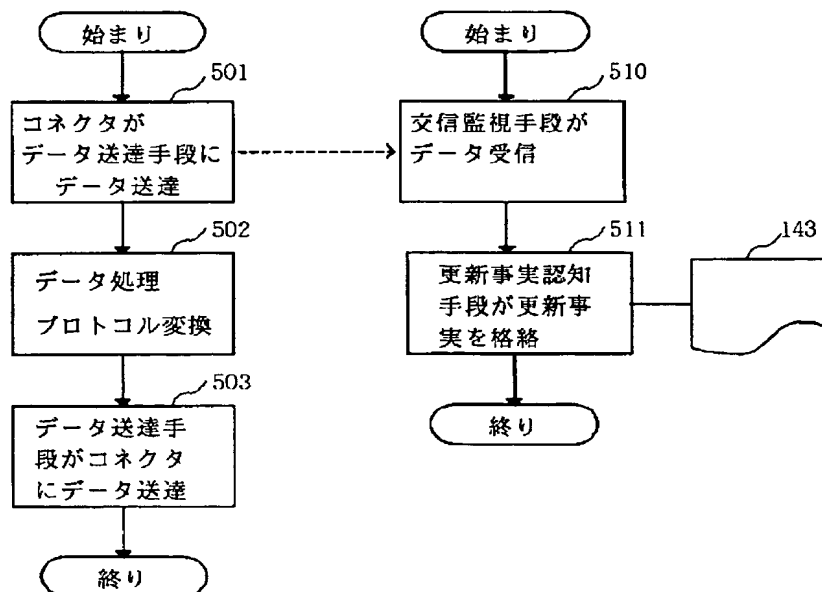
【図3】

【図3】



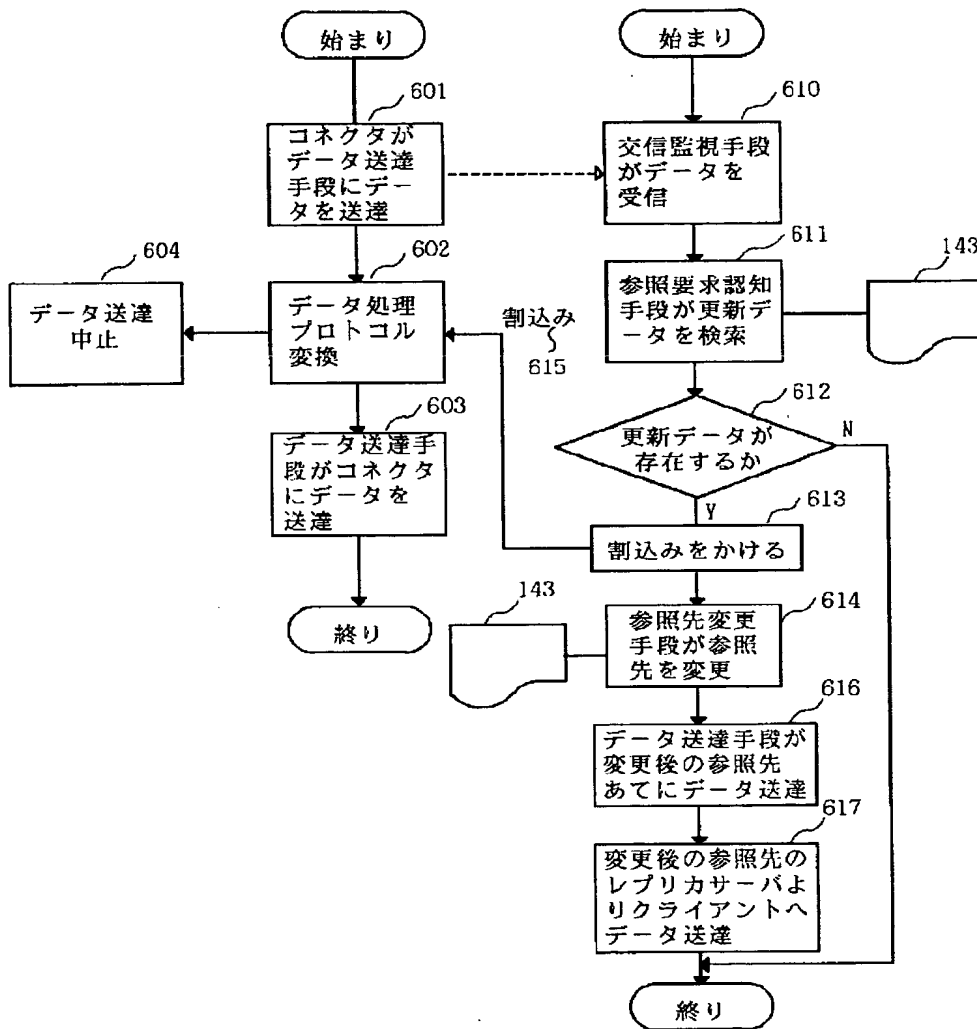
【図5】

【図5】



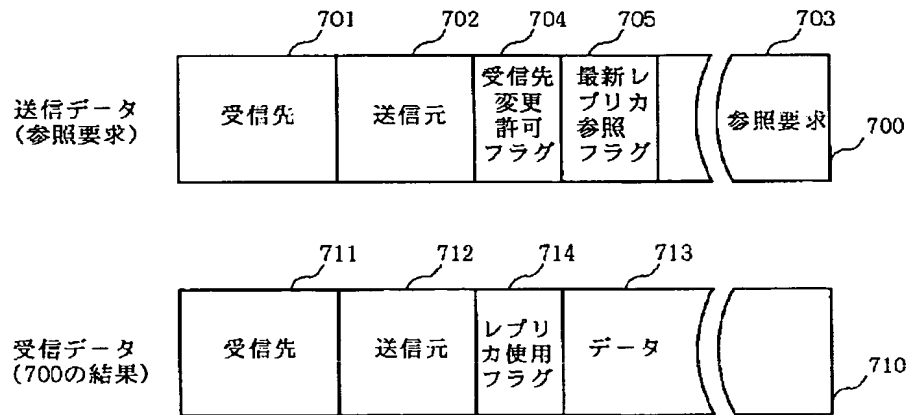
【図6】

【図6】



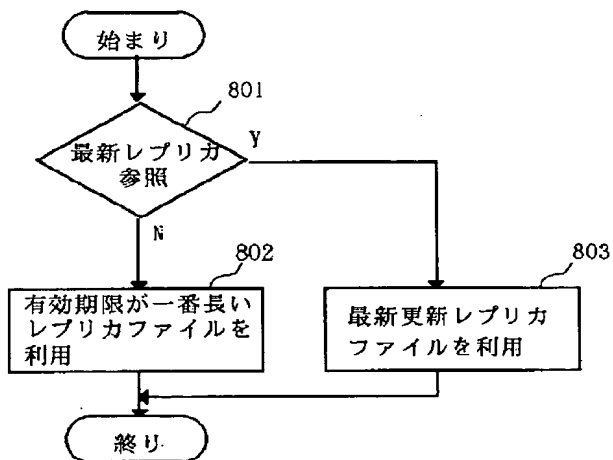
【図7】

【図7】



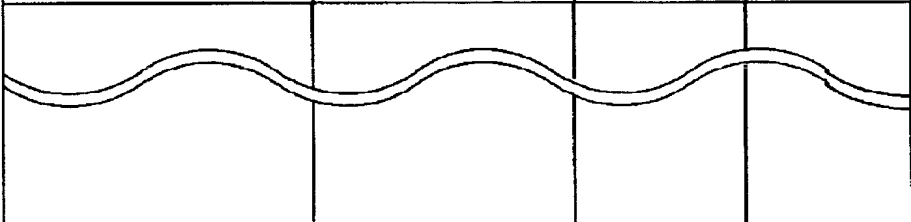
【図8】

【図8】



【図9】

【図9】

オリジナルファイル	レプリカ	有効期限	更新時刻
サーバ110／ファイル112	クライアント 120／122	〇〇〇〇	××××
サーバ110／ファイル112	クライアント △△△／×××	□□□□	△△△△
			

フロントページの続き

(72)発明者 中野 裕彦
神奈川県川崎市麻生区王禅寺1099番地 株
式会社日立製作所システム開発研究所内

(72)発明者 小林 敦
神奈川県横浜市戸塚区戸塚町5030番地 株
式会社日立製作所ソフトウェア開発本部内